Smoothing under diffeomorphic constraints with homeomorphic splines

Jérémie Bigot & Sébastien Gadat
Institut de Mathématiques de Toulouse
Université Paul Sabatier
F-31062 Toulouse Cedex 9, France
email: {Jeremie.Bigot,Sebastien.Gadat}@math.univ-toulouse.fr

December 7, 2009

Abstract

In this paper we introduce a new class of diffeomorphic smoothers based on general spline smoothing techniques and on the use of some tools that have been recently developed in the context of image warping to compute smooth diffeomorphisms. This diffeomorphic spline is defined as the solution of an ordinary differential equation governed by an appropriate time-dependent vector field. This solution has a closed form expression which can be computed using classical unconstrained spline smoothing techniques. This method does not require the use of quadratic or linear programming under inequality constraints and has therefore a low computational cost. In a one dimensional setting incorporating diffeomorphic constraints is equivalent to impose monotonicity. Thus, as an illustration, it is shown that such a monotone spline can be used to monotonize any unconstrained estimator of a regression function, and that this monotone smoother inherits the convergence properties of the unconstrained estimator. Some numerical experiments are proposed to illustrate its finite sample performances, and to compare them with another monotone estimator. We also provide a two-dimensional application on the computation of diffeomorphisms for landmark and image matching.

Key words and phrases: Reproducing Kernel Hilbert Space, Constrained smoothing; Monotonicity; Splines; Nonparametric regression; Diffeomorphism; Ordinary differential equation; Time-dependent vector field.

AMS 1991 subject classifications: Primary 62G08; secondary 65Dxx.

1 Introduction

Spline smoothing is widely used in many different areas to study the relationship between a response variable Y and an independent variable X, see e.g. Wahba [42] for a detailed presentation, and have many applications in approximation problems, see e.g. Duchon [15], Quak & Schumaker [33], de Boor & Schumaker [12] or Lopez de Silanez & Apprato [26] . In many fields of interest, including physical and medical sciences, one is often interested in imposing a monotonic relationship between two such variables. Typical examples

include the analysis of dose-response curves in pharmakinetics (Kelly & Rice [24]), growth curves in biology and many specific practical problems discussed in the literature cited below. Note that without loss of generality, monotone smoothing is considered in this paper as the problem of computing an increasing function. For calculating a decreasing smoother one can simply reverse the "X axis" and then apply the same methodology.

Unconstrained spline smoothing consists in minimizing over an appropriate functional space $\mathcal{F}$ a criterion that represents a balance between two terms : fidelity to the data and smoothness of the fitting spline. If $\mathcal{F}$ is a Reproducing Kernel Hilbert Spaces (RKHS), then it is well known (see e.g. Wahba [42]) that a closed form solution can be computed by solving a simple linear system of equations. The simplest idea that comes to mind to incorporate monotonicity constraints is to restrict the search space $\mathcal{F}$ to a subset of monotone functions, and then to take as a monotone smoother the function which minimizes the same criterion over this restricted space. Existence, approximation properties and the numerical computation of such monotone smoothers have been widely studied, see e.g. Utreras [40], Andersson & Elfving [1], Elfving & Andersson [16], Irvine, Marin & Smith [23], Beatson & Ziegler [7].

However, finding the exact solution of a smoothing problem in a constrained space is generally a difficult task and most existing algorithms only produce approximate solutions. As a closed form solution for such monotone smoothers does not exist in general, their numerical computation is generally done by determining the fitted values of the smoothing spline on a finite set of points (usually the observed covariates) and uses a set of inequality constraints to impose restrictions on the value of the fitted function at these points. However, the algorithms used to compute these estimators can be computationally intensive since they involve a large set of inequality constraints (see e.g. Schwetlick & Kunert [35], He & Shi [22], Turlach [39] and the discussion therein).

In this paper, we introduce a new class of monotone smoothers that have a closed form expression which depends on the underlying RKHS, and which can be computed using classical unconstrained spline smoothing techniques. Thus, unlike some monotone smoothers, this method does not require the use of quadratic programming under linear constraints and has therefore a low computational cost. Our approach is based on tools that have been recently developed in the context of image warping for the construction of diffeomorphisms in two or three dimensions (see e.g. Trouvé [38], Miller, Trouvé & Younes [30], Glaunès [19], Younes [43], [44], see also Apprato & Gout [2] for the use of diffeomorphism in spline approximation). Trouvé, Younes and their collaborators have proposed to compute smooth diffeomorphisms as the solutions of a new class of ordinary differential equation governed by a time-dependent velocity field. In a one-dimensional (1D) setting, it is easy to see that a diffeomorphism is a smooth and monotone function, and thus the main idea of this paper is to adapt such tools for the construction of 1D monotone smoothers. Our approach also yields a new method to compute smooth diffeomorphisms for the alignment of landmarks in a 2D or 3D setting. Some examples in the 2D case for image warping (see e.g. Bigot *et al.* [9]) are given in the section on numerical experiments, but for simplicity our theoretical results are presented in a 1D setting.

Our main contributions are the following: first we show how one can generate a strictly monotone function as the solution of an ordinary differential equation (ODE). We also prove that for some functional classes, any monotone function can be represented as the solution of such an ODE. Secondly, a new criterion to fit a monotone spline is proposed and we show that the minimizer of this criterion has a simple closed form expression. As an illustrative example, we explain how the overall methodology can be applied to the problem of monotonizing any estimator of a regression function. Indeed, in statistics,

a possible smoothing method under shape constraints consists in first using an unconstrained estimator (such as a spline, wavelets or kernel smoother) and then projecting the resulting curve estimate onto a constrained subspace of regression functions which is usually a convex set (see e.g. Mammen, Marron, Turlach & Wand [28] and Mammen & Thomas-Agnan [29]). For the problem of monotone regression, this approach is generally referred to as smooth and then monotonize. However, as pointed out by Gijbels [18] many of these monotone estimates appear less smooth than the unconstrained estimates due to the projection step. Moreover, it is not clear how one can compute numerically the projection of a curve estimate onto a constrained subspace for any unconstrained estimator. Our monotone estimator does not suffer from these two drawbacks since it can be easily computed, and it yields surprisingly very smooth estimates.

The remainder of the paper is structured as follows: section 2 gives a brief overview of RKHS and the general spline smoothing problem. In section 3, we show how one can generate a strictly monotone function as the solution of an ODE. In section 4, we propose a new class of monotone smoothers that we call homeomorphic smoothing splines. In Section 5, we apply this methodology to nonparametric regression under monotonicity constraints. Section 6 presents a short Monte Carlo study on the efficiency of this approach and a comparison with another constrained estimate. Finally, another application of homeomorphic splines is presented in a 2D setting for matching problems involving the alignment of landmarks. The Appendix provides the proofs of the main results.

2 The general spline smoothing problem in the 1D case

Let $\mathcal{H}_K$ a be RKHS of functions in $L^2(\mathbb{R})$ with positive definite kernel K, meaning that for all $x \in \mathbb{R}$ there exists an element $K_x(.)$ such that for all g in $\mathcal{H}_K$, we have $g(x) = \langle K_x(.), g \rangle_K$ where $\langle ., . \rangle_K$ denotes the scalar product on $\mathcal{H}_K$ whose derived norm is $\|.\|_K$ (for more details on RKHS we refer to Atteia [5], Atteia & Gaches [6], Aronszajn [4], Duchon [15], Wahba [42], Berlinet & Thomas-Agnan [8]). Let $\psi_1, \ldots, \psi_M$ be functions (not necessarily in $L^2(\mathbb{R})$) such that for any set of distinct points $x_1, \ldots, x_n$ in $\mathbb{R}$, the matrix T with elements $T_{i,j} = \psi_j(x_j)$ has full rank $M < n$. Let $\tilde{\mathcal{H}} = Span\{\psi_j\}_{j=1,\ldots,M} + \mathcal{H}_K$ and assume that we have n distinct pairs of points $(x_i, y_i) \in \mathbb{R}^2$. Then the general spline smoothing problem is to find the minimizer

$$\tilde{h}_{n,\lambda} = \arg\min_{\tilde{h} \in \tilde{\mathcal{H}}} \frac{1}{n} \sum_{i=1}^{n} (\tilde{h}(x_i) - y_i)^2 + \lambda \|h\|_K^2, \qquad (2.1)$$

for any $\tilde{h} \in \tilde{\mathcal{H}}$ of the form $\tilde{h}(x) = \sum_{j=1}^{M} a_j \psi_j(x) + h(x)$, where $h \in \mathcal{H}_K$. It is well known that the solution of this smoothing problem is unique and of the form: $\forall x \in \mathbb{R}$, $\tilde{h}_{n,\lambda}(x) = \sum_{j=1}^{M} \alpha_j \psi_j(x) + \sum_{i=1}^{n} \beta_i K(x, x_i)$, where α, β are solutions of a simple linear system of equations (see e.g. Wahba [42]).

Throughout this paper, we will assume $M = 2$ with $\psi_1(x) = 1$ and $\psi_2(x) = x$, these conditions are convenient for satisfying the uniform Lipschitz condition stated in Lemma 7.1 (see the Appendix).

As an example of a RKHS, we will often use the Sobolev space of order $m \in \mathbb{N}^*$ (see Berlinet and Thomas-Agnan [8]) $\mathcal{H}_K = \mathcal{H}^m(\mathbb{R})$ endowed with the norm $\|h\|_{\mathcal{H}^m}^2 = \int_{-\infty}^{+\infty} |h(x)|^2 dx + \int_{-\infty}^{+\infty} |h^{(m)}(x)|^2 dx$. With this choice for the ψ_j's, any function $\tilde{h} \in \tilde{\mathcal{H}}$ is of

the form $\tilde{h}(x) = a_1 + a_2 x + h(x)$ where $a_1, a_2 \in \mathbb{R}$ and $h \in \mathcal{H}_K$. Hence, one can define a norm in $\tilde{\mathcal{H}}$ by setting

$$\|\tilde{h}\|_{\tilde{\mathcal{H}}} = \max(|a_1|, |a_2|) + \|h\|_K.$$

Note that $\tilde{\mathcal{H}}$ is a Banach space for this norm. From now, to simplify the notations we will omit the superscript $\tilde{\mathcal{H}}$ and write $\|\tilde{h}\|_{\tilde{\mathcal{H}}} = \|\tilde{h}\|$.

3 Differential Equation to generate monotone functions

The smoothing spline $\tilde{h}_{n,\lambda}$ defined previously is not necessarily a monotone function. We thus propose to use a connection between monotone functions and time-dependent vector fields to incorporate monotonicity constraints into the computation of $\tilde{h}_{n,\lambda}$.

3.1 Generating monotone functions

Let us explain the basic ideas (as described e.g. in Younes [43]) to generate smooth diffeomorphisms. Take any $v \in \mathcal{C}^1(\mathbb{R}, \mathbb{R})$ with $\|v'\|_\infty < +\infty$. Then if $\epsilon > 0$ is chosen sufficiently small the perturbation $\phi = Id + \epsilon v$ of the identity function, is a strictly increasing *small* diffeomorphism. Now, if $v_{t_1}, \ldots, v_{t_p}$ are continuously differentiable functions on $\mathbb{R}$ and if $\epsilon > 0$ is sufficiently small such that $Id + \epsilon v_{t_k}$ are small diffeomorphisms on $\mathbb{R}$, then we can construct the following sequence of diffeomorphisms $\phi_{t_p} = (Id + \epsilon v_{t_{p-1}}) \circ \ldots \circ (Id + \epsilon v_{t_1})$. Then, note that $\phi_{t_{p+1}} = (Id + \epsilon v_{t_p}) \circ \phi_{t_p} = \phi_{t_p} + \epsilon v_{t_p} \circ \phi_{t_p}$ which can also be written as

$$\forall x \in \mathbb{R} \qquad \frac{\phi_{t_{p+1}}(x) - \phi_{t_p}(x)}{\epsilon} = v_{t_p}\left[\phi_{t_p}(x)\right]. \tag{3.1}$$

As $\epsilon \to 0$ and $\lim_{p \to +\infty} t_{p+1} - t_p = 0$, (3.1) looks like a discretized version of an inhomogeneous differential equation of the following form (by introducing a continuous time variable t):

$$\frac{\mathrm{d}\phi_t}{\mathrm{d}t} = v_t(\phi_t). \tag{3.2}$$

In the sequel, v will be a function of two variables (t, x), and for a fixed t we will use the notation $x \mapsto v_t(x)$ to refer to the application $x \mapsto v(t, x)$. The variable t varies in the finite time interval $[0; 1]$ while x belongs to $\mathbb{R}$. Similarly ϕ depends both on the time t and the variable x, and $\phi_t(x)$ will refer to $\phi(t, x)$. Thus, equation (3.2) is equivalent to: $\forall x \in \mathbb{R}, \frac{\mathrm{d}\phi(t,x)}{\mathrm{d}t} = v(t, \phi(t, x))$.

As we will see, under mild conditions on the time-dependent vector field $(v_t)_{t \in [0,1]}$, the solution of the above ODE is a diffeomorphism at all time t and thus a monotone function. The main idea of this paper is thus the following: we transfer the problem of computing a monotone spline from a set of n data points $(x_i, y_i)_{i=1\ldots n}$ to the problem of computing an appropriate vector field $(v_t^n)_{t \in [0,1]}$ which depends on these data points. A monotone smoother f_n is then defined as the solution at time $t = 1$ of the ODE (3.2) governed by the vector field $(v_t^n)_{t \in [0,1]}$, i.e $f_n(x) = \phi_1(x)$ with $\phi_0(x) = x$. The main advantage of this approach is that the computation of $(v_t^n)_{t \in [0,1]}$ will be obtained from an unconstrained smoothing problem, and therefore the calculation of f_n only requires to run an ODE without imposing any specific constraints. This yields a fitting function f_n which is guaranteed to be monotone.

3.2 Vector fields and ODE

Following the notations in Younes [44], let us state several definitions.

Definition 3.1 ($\mathcal{X}^1, \mathcal{X}^2$ and $\mathcal{X}$) *$\mathcal{X}^1$ is the space of time-dependent vector fields ($v_t \in \tilde{\mathcal{H}}, t \in [0,1]$) such that $\|v\|_{\mathcal{X}^1} =_{def} \int_0^1 \|v_t\| dt < +\infty$. $\mathcal{X}^2$ is the space of time-dependent vector fields ($v_t \in \tilde{\mathcal{H}}, t \in [0,1]$) such that $\|v\|_{\mathcal{X}^2} =_{def} \int_0^1 \|v_t\|^2 dt < +\infty$. Finally, $\mathcal{X}$ is the set of all time-dependent vector field ($v_t \in \tilde{\mathcal{H}}, t \in [0,1]$).*

The definitions of $\mathcal{X}, \mathcal{X}^1, \mathcal{X}^2$ are classical in the context of time-dependent PDE which are formulated as Banach space-valued functions, see e.g. Renardy & Rogers [34]. Note that by the Cauchy-Schwarz inequality, $\|v\|_{\mathcal{X}^1}^2 \leq \|v\|_{\mathcal{X}^2}$, and thus $\mathcal{X}^2 \subset \mathcal{X}^1 \subset \mathcal{X}$. For $v \in \mathcal{X}^1$, we formally define an ODE governed by the time-dependent vector field ($v_t, t \in [0,1]$) as

$$\frac{d\phi_t}{dt} = v_t(\phi_t). \tag{3.3}$$

Definition 3.2 *Let $\Omega = [0,1]$. A function $t \mapsto \phi_t$ is called a solution of the equation (3.3) with initial condition the identity if for all $x \in \Omega$, $t \mapsto \phi_t(x)$ is a continuous function from $[0,1]$ to $\mathbb{R}$, $\phi_0(x) = x$ for all $x \in \Omega$, and for all $t \in [0,1]$ and all $x \in \Omega$, $\phi_t(x) = x + \int_0^t v_s(\phi_s(x)) ds$.*

The following theorem, whose proof is deferred to the appendix, shows that the solution of the equation (3.3) is unique and is a homeomorphism for all time $t \in [0,1]$.

Theorem 3.1 *Assume that the kernel K is bounded on $\mathbb{R}^2$, and that there exists a constant C_1 such that for any $h \in \mathcal{H}_K$*

$$|h(x) - h(y)| \leq C_1 \|h\|_K |x - y|. \tag{3.4}$$

Let $v \in \mathcal{X}^1$. Then, for all $x \in \Omega$ and $t \in [0,1]$, there exists a unique solution of (3.3) with initial condition the identity. Besides, for all $t \in [0,1]$, ϕ_t is a homeomorphism from Ω to $\phi_t(\Omega)$.

The above uniformly Lipschitz assumption (3.4) for the kernel K is not restrictive as it is satisfied in many cases of interest. Indeed, observe that for for any $h \in \mathcal{H}_K$:

$$|h(x) - h(y)| = |\langle K(.,x) - K(.,y), h \rangle_{\mathcal{H}_K}| \leq \|h\|_K \|K(.,x) - K(.,y)\|_K.$$

If K is a radial kernel of the form $K(x,y) = k(|x-y|)$ for some function $k : \mathbb{R} \to \mathbb{R}$, then the above equation implies that $|h(x) - h(y)| \leq 2\|h\|_K |k(0) - k(|x-y|)|$. Hence, h satisfies equation (3.4) provided k is uniformly Lipschitz on $\mathbb{R}$. This is the case for a Gaussian kernel: $k(|x-y|) = e^{-|x-y|^2/2\sigma^2}$, and also in the Sobolev case where K is given by (see e.g. [8]): $K(x,y) = k_m(|x-y|) := \sum_{k=0}^{m-1} \frac{\exp\left[-|x-y|\exp(i\frac{\pi}{2m} + \frac{k\pi}{m} - \frac{\pi}{2})\right]}{2m\exp((2m-1)(i\frac{\pi}{2m} + \frac{k\pi}{m}))}$.

Remark: this framework can be extended to a 2D setting for generating diffeomorphism of $\mathbb{R}^2$. For this, let $\tilde{\mathcal{H}}_2$ denote a set of smooth functions from $\mathbb{R}^2$ to $\mathbb{R}^2$ (see Section 6.2 for an example) and define $\tilde{\mathcal{X}} = \{(v_t, t \in [0,1]) \text{ with } v_t \in \tilde{\mathcal{H}}_2 \text{ for all } t \in [0,1]\}$. Let Let Ω be an open subset of $\mathbb{R}^2$ and define an ODE governed by the time-dependent vector field $(v_t, t \in [0,1]) \in \tilde{\mathcal{X}}$ as: $\frac{d\phi_t^v}{dt} = v_t(\phi_t^v)$, with $v_0(x) = x$. Using arguments in the proof of Theorem 3.1 and assuming that the functions h in $\tilde{\mathcal{H}}_2$ are sufficiently smooth and satisfy a uniform Lipschitz condition of the type (3.4), then one can easily show that the solution of such an ODE is unique and is a diffeomorphism from Ω to $\phi_t(\Omega) \subset \mathbb{R}^2$ for all time $t \in [0,1]$.

4 Homeomorphic smoothing splines

4.1 A connection between monotone functions and time-dependent vector fields

A natural question is to ask if any monotone function can be written as the solution of an ODE governed by a time-dependent vector field. First, consider the case where $\mathcal{H}_K = \mathcal{H}^m(\mathbb{R})$ and f belongs to the Sobolev space

$$H^m([0,1]) = \{f : [0,1] \to \mathbb{R}, f^{(m-1)} \text{ is absolutely continuous with } \int_0^1 |f^{(m)}(x)|^2 dx < +\infty\}.$$

Then, if f is monotone, one of our main results is the following theorem which states that f can be represented as the solution at time $t = 1$ of an ODE:

Theorem 4.1 *Assume that $\tilde{\mathcal{H}} = Span\{1, x\} + \mathcal{H}^m(\mathbb{R})$. Let $m \geq 2$ and $f \in H^m([0,1])$ be such that $f'(x) > 0$ for all $x \in [0,1]$ and define $\phi_t(x) = tf(x) + (1-t)x$, for all $t \in [0,1]$. Then, there exits a time-dependent vector field $(v_t^f)_{t \in [0,1]}$ depending on f, such that $v_t^f \in \mathcal{H}^m(\mathbb{R})$ for all $t \in [0,1]$ and which satisfies $\phi_1 = \phi_0 + \int_0^1 v_t^f(\phi_t)dt$, and thus $f(x) = \phi_1(x) = x + \int_0^1 v_t^f(\phi_t(x))dt$. Moreover for all $t \in [0,1]$ one has that*

$$v_t^f(\phi_t(x)) = v_t^f(tf(x) + (1-t)x) = f(x) - x = \frac{d\phi_t}{dt}(x) \text{ for all } x \in [0,1]. \qquad (4.1)$$

For all $t \in [0,1]$, the function v_t^f can be chosen as the unique element of minimal norm in $H^m(\mathbb{R})$ which satisfies equation (4.1).

To the best of our knowledge, this representation of a monotone function by such an ODE has not been used before. The formulation (4.1) suggests the following trick to compute a monotone smoother from a set of n data points $(x_i, y_i) \in [0,1] \times \mathbb{R}$: if one considers y_i as an approximation of $f(x_i)$ for some function f, then to obtain a good approximation of v^f, one can use the y_i's to compute a vector field v^n that satisfies roughly the interpolating conditions (4.1) at the design points. More precisely, at any time t, the vector field v_t^n is obtained by smoothing the "data" $(ty_i + (1-t)x_i, y_i - x_i)$, $i = 1, \ldots, n$. Finally, to compute a monotone smoother f_n we just have to run the ODE (3.3) with the vector field v^n.

When $\mathcal{H}_K \neq \mathcal{H}^m(\mathbb{R})$, it is not clear if one can obtain a general correspondence between monotone functions and their representation via a vector field $v \in \mathcal{X}^2$. However, we believe that the proof of Theorem 4.1 could be adapted to other RKHS.

4.2 A new monotone smoothing spline

Let $(x_i, y_i), i = 1, \ldots, n$ be a set of data points with $x_i \in [0, 1]$ and $y_i \in \mathbb{R}$. A new smoothing spline problem under monotonicity constraints can be formulated in the following way: for a time-dependent vector field $v \in \mathcal{X}$, define the "energy"

$$E_\lambda(v) = \int_0^1 \frac{1}{n} \sum_{i=1}^n \left(y_i - x_i - v_t(ty_i + (1-t)x_i)\right)^2 dt + \lambda \int_0^1 \|h_t\|_K^2 dt, \qquad (4.2)$$

where $v_t(x) = a_1^t + a_2^t x + h_t(x)$, and $\lambda > 0$ is a regularization parameter. Then, take $v^{n,\lambda} = \arg\min_{v \in \mathcal{X}} E_\lambda(v)$, and a monotone smoother $f_{n,\lambda}^c$ is obtained by taking $f_{n,\lambda}^c(x) = \phi_1^{v^{n,\lambda}}(x) = x + \int_0^1 v_t^{n,\lambda}(\phi_t^{v^{n,\lambda}}(x)) dt$. The following proposition gives sufficient conditions for the existence of $v^{n,\lambda}$:

Proposition 4.1 *Assume that the conditions of Theorem 3.1 are satisfied. Assume that $n > 2$ and that the kernel $K : \mathbb{R}^2 \to \mathbb{R}$ is continuous. Suppose that the x_i's and the y_i's are such that the n "design points" $ty_i + (1-t)x_i$ are distinct in $\mathbb{R}$ for any $t \in [0, 1]$. Then, the optimization problem (4.2) has a unique solution $v^{n,\lambda} \in \mathcal{X}$ such that at each time $t \in [0, 1]$, $v_t^{n,\lambda}$ is the solution of the following standard unconstrained smoothing problem: find $v_t \in \tilde{\mathcal{H}}$ which minimizes*

$$E_\lambda^t(v_t) = \frac{1}{n} \sum_{i=1}^n \left(y_i - x_i - v_t(ty_i + (1-t)x_i)\right)^2 + \lambda \|h_t\|_K^2, \qquad (4.3)$$

where $v_t(x) = a_1^t + a_2^t x + h_t(x)$. Moreover $v^{n,\lambda} \in \mathcal{X}^2$, and $f_{n,\lambda}^c$ is a monotone function on $[0, 1]$.

Let us remark that if one defines $\tilde{\mathcal{H}}^c$ as the subspace of functions $f \in \tilde{\mathcal{H}}$ such that f is a strictly monotone function on $[0, 1]$, then a monotone smoother $\tilde{f}_{n,\lambda}^c$ can de defined by mininizing the classical spline smoothing criterion over the restricted space $\tilde{\mathcal{H}}^c$ i.e.

$$\tilde{f}_{n,\lambda}^c = \arg\min_{\tilde{h} \in \tilde{\mathcal{H}}^c} \frac{1}{n} \sum_{i=1}^n (\tilde{h}(x_i) - y_i)^2 + \lambda \|h\|_K^2. \qquad (4.4)$$

General smoothing splines problems under under shape constraints such as monotonicity have been studied in detail in Utreras [40]. Theorems proving the existence, uniqueness and general results concerning the characterization of $\tilde{f}_{n,\lambda}^c$ are given in Utreras [40], together with a study of the convergence rate of $\tilde{f}_{n,\lambda}^c$ in a nonparametric regression setting. Hence, it would be interesting to study the relationship that may exist between the estimators $\tilde{f}_{n,\lambda}^c$ and $f_{n,\lambda}^c$. However, we believe that this problem is not an easy task which is beyond the scope of this paper.

4.3 Computational aspects and the choice of λ_n

Numerical computation   The optimization problem (4.3) amounts to solve, at each time $t \in [0, 1]$, a simple finite-dimensional least-square problem which yields a very simple algorithm to compute a smooth increasing function: choose a discretization $t_k = \frac{k}{T}, k = 0, \ldots, T - 1$ of the time-interval $[0, 1]$ (we took $T = 30$) and set $\phi_{n,\lambda}^{t_0}(x) = x$ for $x \in [0, 1]$.

Then repeat for $k = 0, \ldots, T - 1$: find the solution $v_{t_k}^{n,\lambda}$ of the unconstrained smoothing problem (4.3) for each $t = t_k$, and then compute $\phi_{n,\lambda}^{t_{k+1}}(x) = \phi_{n,\lambda}^{t_k}(x) + \frac{1}{T} v_{t_k}^{n,\lambda}\left(\phi_{n,\lambda}^{t_k}(x)\right)$. The proposed numerical scheme is based on

$$\phi_{n,\lambda}^{t_{k+1}} = (Id + \frac{1}{T} v_{t_k}^{n,\lambda})(\phi_{n,\lambda}^{t_k}) \tag{4.5}$$

which replaces the theoretical relation $\phi_{n,\lambda}^{t_{k+1}} = \phi_{n,\lambda}^{t_k} + \int_{t_k}^{t_{k+1}} v_u^{n,\lambda}(\phi_{n,\lambda}^u) du$. Remark that equation (4.5) shows that if $\frac{\|v_t^{n,\lambda}\|}{T} < 1$ for all t, then $\phi_{n,\lambda}^{t_{k+1}}$ remains monotone provided $\phi_{n,\lambda}^{t_k}$ is monotone. This condition is not really restrictive since we have shown that $(v_t^{n,\lambda})_{t \in [0,1]}$ is in $\mathcal{X}^2$ and that $t \to v_t^{n,\lambda}$ is a continuous map on $[0;1]$. Thus, our estimator based on the Euler scheme (4.5) remains monotone if T is chosen sufficiently large, namely greater than $\sup_{t \in [0,1]} \|v_t^{n,\lambda}\|$.

Another important question is the error made using the Euler discretization scheme instead of the correct ODE This point is left open since it is far beyond the scope of this paper but the use of the Gronwall Lemma should enable to derive upper bound between the theoretical $\phi_{n,\lambda}^1$ and the approximated one derived from (4.5).

Choice of the regularization parameter A fundamental issue is the choice of the regularization parameter λ. In our simulations, we have obtained good results via an empirical choice of λ inspired by the generalized cross-validation (GCV) criterion of Craven & Wahba [11], see also Girard [20] for fast cross-validation methods. For $i = 1, \ldots, n$ and $t \in [0, 1]$, define $\hat{X}_i^t = ty_i + (1 - t)x_i$ and $\hat{Y}_i = y_i - x_i$. Then, note that at each time $t \in [0, 1]$ the smoothing spline $v_t^{n,\lambda}$ evaluated at the "design points" $\hat{X}_i^t, \ldots, \hat{X}_n^t$ is a linear function of the observations $\hat{Y}_1, \ldots, \hat{Y}_n$, i.e. there exists a matrix $A_{\lambda,t}$ such that $\mathbf{v}_t^{n,\lambda} = \left(v_t^{n,\lambda}(\hat{X}_1^t), \ldots, v_t^{n,\lambda}(\hat{X}_n^t)\right)' = A_{\lambda,t}\mathbf{Y}$, with $\mathbf{Y} = (\hat{Y}_1, \ldots, \hat{Y}_n)'$. Therefore, to choose the smoothing parameter λ, we simply propose to minimize the following empirical GCV-type criterion :

$$V(\lambda) = \frac{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{f}_{n,\lambda}^c(x_i))^2}{\int_0^1 [Tr(I_n - A_{\lambda,t})]^2 dt}. \tag{4.6}$$

In the above equation, the quantity $\int_0^1 [Tr(I_n - A_{\lambda,t})]^2 dt$ can be interpreted as a measure of the degree of freedom of the smoothing spline $\hat{f}_{n,\lambda}^c$. The quantity $V(\lambda)$ is therefore the classical GCV criteria which is the ratio between the empirical error and the complexity of a smoothing procedure. To set a good penalization parameter λ, we simply use a grid search to minimize $V(\lambda)$.

Computational cost The proposed method has a relatively low computational cost compared to classical constrained optimization methods. If $t_0, \ldots, t_{T-1}$ denotes a discretization of $[0, 1]$ (with T ndependent of n), our method requires for each t_k the inversion of a symmetric definite matrix of size $n \times n$ which is possible using $\mathcal{O}(n^3)$ operations with a Cholesky algorithm for instance. The computational cost of our method is thus $\mathcal{O}(Tn^3)$. Numerical computation of a constrained spline smoothing problem such as (4.4) is generally done by using a set of inequality constraints to impose monotonicity on the value of the fitted function at a finite number of points. However, such algorithms can be computationally intensive since solving a general problem of quadratic optimization with linear constraints is generally NP-hard (see e.g. Pardalos & Vavasis [32]). Primal-dual methods for instance can iteratively solve the problem but their complexity is larger than $\mathcal{O}(n^3)$.

5 A non-parametric regression problem under monotonicity constraints

Consider the standard nonparametric regression problem on a bounded interval:

$$y_i = f(x_i) + \epsilon_i, \ i = 1, \dots, n, \tag{5.1}$$

where $f : [0,1] \to \mathbb{R}$ and ϵ_i are independent and identically distributed (i.i.d.) variables with zero mean and variance σ^2. The regression function f is assumed to belong to a class of strictly increasing functions $\mathcal{F}$ that satisfy some smoothness conditions to be defined later. Smoothing procedures for monotone regression can be found in He & Shi [22], Kelly & Rice [24], Mammen [27], Mammen & Thomas-Agnan [29], Hall & Huang [21], Mammen, Marron, Turlach & Wand [28], Dette, Neumeyer & Pilz [13] and Antoniadis, Bigot & Gijbels [3].

In this section, we explain how homeomorphic splines can be used as a smooth and then monotonize method. Let $\hat{f}_n$ be an unconstrained estimator obtained from the data $(y_i, x_i), i = 1, \dots, n$ (e.g. by spline, kernel or wavelet smoothing). Our goal is to construct a monotone estimator $\hat{f}_n^c$ which inherits the asymptotic properties of the unconstrained estimator $\hat{f}_n$ in terms of the empirical mean squared error: $R_n(\hat{f}_n^c, f) = \frac{1}{n} \sum_{i=1}^n (\hat{f}_n^c(x_i) - f(x_i))^2$. For this, starting from the values $\hat{f}_n(x_i)$ instead of the observed y_i's, take the vector field $v^{n,\lambda}$ which minimizes the following criterion:

$$v^{n,\lambda} = \arg \min_{v \in \mathcal{X}} \int_0^1 \frac{1}{n} \sum_{i=1}^n \left(\hat{f}^n(x_i) - x_i - v_t(t \hat{f}^n(x_i) + (1-t)x_i) \right)^2 dt + \lambda \int_0^1 \|h_t\|_K^2 dt,$$

where $v_t(x) = a_1^t + a_2^t x + h_t(x)$. Then $\hat{f}_{n,\lambda}^c$ is defined as the solution at time $t = 1$ of the ODE (3.3) governed by the time-dependent vector field $v^{n,\lambda}$.

5.1 Asymptotic properties of the monotone estimator $\hat{f}_{n,\lambda}^c$

The following theorem shows that under mild conditions on the design and the unconstrained estimator, the monotone estimator $\hat{f}_{n,\lambda}^c$ inherits the asymptotic properties of $\hat{f}^n$ in term of rate of convergence. To the best of our knowledge, this is the first consistency result on estimators defined through large diffeomorphism models governed by ODE.

Theorem 5.1 *Assume that the conditions of Theorem 3.1 are satisfied, and that the kernel $K : \mathbb{R}^2 \to \mathbb{R}$ is continuous. Moreover assume that the function f is continuously differentiable on $[0,1]$ with $f'(x) > 0$ for all $x \in [0,1]$ and that there exists a time-dependent vector field $v^f \in \mathcal{X}^2$ such that for all $t \in [0,1]$:*

$$v_t^f(\phi_t(x)) = f(x) - x \text{ for all } x \in [0,1],$$

where $\phi_t(x) = tf(x) + (1-t)x$. Suppose that the unconstrained estimator $\hat{f}^n$ and the points $x_i, i = 1, \dots, n$ satisfy the following property:

A1 *for all $t \in [0,1]$ and all $1 \le i, j \le n$ with $i \ne j$,*

$$t \hat{f}^n(x_i) + (1-t)x_i \ne t \hat{f}^n(x_j) + (1-t)x_j \ a.s. \tag{5.2}$$

Then, $v^{n,\lambda} \in \mathcal{X}^2$ a.s. and thus $\hat{f}^c_{n,\lambda}$ is a monotone function on $[0,1]$. If we further assume that:

A2 *there exists a weight function $\omega : [0,1] \to]0, +\infty[$ such that for any $g \in \mathcal{C}^1([0,1], \mathbb{R})$ one has $\lim_{n \to +\infty} \frac{1}{n} \sum_{i=1}^n g(x_i) \to \int_0^1 g(x) \omega(x) dx$,*

A3 *$R_n(\hat{f}_n, f) \to 0$ in probability as $n \to +\infty$.*

Then, for any sequence $\lambda = \lambda_n \to 0$, we have that there exists a deterministic constant Λ_1 (not depending on n) such that with probability tending to one as $n \to +\infty$:

$$R_n(\hat{f}^c_{n,\lambda_n}, f) \leq \Lambda_1 \left(R_n(\hat{f}_n, f) + \lambda_n \right).$$

Equation (5.2) may not be satisfied for time points t_{ij} such that $\frac{1-t_{ij}}{t_{ij}} = \frac{\hat{f}^n(x_i) - \hat{f}^n(x_j)}{x_j - x_i}$. Since the function $t \to \frac{1-t}{t}$ is injective, this can only happen for a finite number of time points $t \in [0,1]$. Hence assumption **A1** is generally satisfied provided the design points are distinct. Moreover, if equation (5.2) is not satisfied for some points t, one can argue that it is possible to modify the estimator $\hat{f}^n$ without changing its asymptotic properties (by slightly varying e.g. the smoothing parameter used to compute it) such that (5.2) is true for any $t \in [0,1]$ and all $1 \leq i, j \leq n$. Note that under assumption **A1**, Proposition 4.1 implies that $v^{n,\lambda}$ can be easily implemented using unconstrained spline smoothing.

The assumption **A2** means that in some sense the design points are sampled according to the density $\omega(x)$. The assumption **A3** is satisfied whenever the expected empirical mean squared error $\mathbb{E}R_n(\hat{f}_n, f)$ converges to zero as $n \to +\infty$. The fact that $v^{n,\lambda} \in \mathcal{X}^2$ guarantees that $\hat{f}_n$ is a monotone function. Moreover, one can see that if λ_n decays as fast as the empirical error $R_n(\hat{f}_n, f)$ then the estimator $\hat{f}^c_{n,\lambda_n}$ has the same asymptotic convergence rate than the unconstrained estimator $\hat{f}_n$. Similar results for smooth and then monotonize approaches are discussed in Mammen & Thomas-Agnan [29], Mammen, Marron, Turlach & Wand [28]. However, the advantages of our approach over existing methods are the following: it yields a monotone smoother which has a closed form expression and which is guaranteed to be monotone on the whole interval $[0,1]$. Moreover, our approach is very general as it is not restricted to functions belonging to Sobolev spaces.

5.2 Optimal rate of convergence for Sobolev spaces

Let us return to the specific case where $f \in H^m([0,1])$ and $\mathcal{H}_K = \mathcal{H}^m(\mathbb{R})$. The asymptotic properties and optimal rates of convergence (in the minimax sense) of unconstrained estimators for functions belonging to Sobolev spaces has been extensively studied (see e.g. Nussbaum [31], Speckman [36]). The estimator $\hat{f}_{n,\gamma}$ of Speckman [36] is based on the use of the Demmler-Reinsch spline basis and on a smoothing parameter γ (see Speckman [36] for further details). Speckman [36] has shown that for an appropriate choice γ^* then $\mathbb{E}R_n(\hat{f}_{n,\gamma^*}, f) = \mathcal{O}\left(n^{-\frac{2m}{2m+1}} \right)$ if $f \in H^m([0,1])$ which is known to be the minimax rate of convergence for functions belonging to Sobolev balls. This result is based on the assumption that the design points are such that $x_i = G((2i-1)/2n)$ where $G : [0,1] \to [0,1]$ is a continuously differentiable function with $G'(x) \geq c > 0$ for some constant c. Hence, the estimator of Speckman [36] satisfies Assumption **A2** with $\omega(x) = \frac{1}{G'(G^{-1}(x))}$, and one

can check that Assumption **A1** also holds. The following corollary is thus an immediate consequence of Theorem 4.1 and Theorem 5.1:

Corollary 1 *Assume that $\tilde{\mathcal{H}} = Span\{1, x\} + \mathcal{H}^m(\mathbb{R})$. Let $m \geq 2$ and $f \in H^m([0, 1])$ be such that $f'(x) > 0$ for all $x \in [0, 1]$. Then, the monotone estimator $\hat{f}^c_{n,\lambda_n}$ based on the minimax estimator $\hat{f}_{n,\gamma^*}$ of Speckman [36] is such that $R_n(\hat{f}^c_{n,\lambda_n}, f) = \mathcal{O}_P\left(n^{-\frac{2m}{2m+1}}\right)$, provided $\lambda_n = \mathcal{O}(n^{-\frac{2m}{2m+1}})$.*

To obtain an adaptive choice of λ (not depending on the unknown regularity m of f), the above Corollary suggests to take $\lambda_n = \frac{1}{n}$ to have a monotone estimator whose empirical mean squared error decays as fast as $R_n(\hat{f}_n, f)$. This choice may yield satisfactory estimates but in our simulations a data-based choice for λ using a GCV criteria (4.6) gives much better results.

6 Numerical experiments

6.1 1D case and monotonicity

Dette, Neumeyer and Pilz [13] have recently proposed another type of smooth and then monotonize method which combines density and regression estimation with kernel smoothers. This approach has been shown to be very successful on many simulated and real data sets (see Dette and Pilz [14]) and we shall therefore use it as a benchmark to assess the quality of our monotone estimator. Similarly to our approach, it requires a preliminary unconstrained estimator $\hat{f}_n$. This estimator is then used to estimate the inverse f^{-1} of the regression function. For this Dette, Neumeyer and Pilz [13] propose to use the following estimator

$$\hat{m}_n^{-1}(x) = \frac{1}{Nh_d} \sum_{i=1}^{N} \int_{-\infty}^{t} K_d\left(\frac{\hat{f}_n(\frac{i}{N}) - u}{h_d}\right) du,$$

where K_d is a positive kernel function with compact support, h_d a bandwidth that controls the smoothness of $\hat{m}_n^{-1}$ and N is an integer not necessarily equal to the sample size n which controls the numerical precision of the procedure. A monotone estimator $\hat{m}_n$ is then obtained by reflection of the function $\hat{m}_n^{-1}$ at the line $y = x$ (see Dette, Neumeyer and Pilz [13] for further details). In Dette and Pilz [13], it is proposed to use a local linear estimate (see Fan and Gijbels [17]) with Epanechnikov kernel for the unconstrained estimator $\hat{f}_n$. The bandwidth h_r of this unconstrained estimate is chosen as $\hat{h}_r = \left(\frac{\hat{\sigma}^2}{n}\right)^{1/5}$, where $\hat{\sigma}^2 = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} \left(y_{(i+1)} - y_{(i)}\right)^2$. For the choice of the bandwidth h_d, it is recommended to choose $h_d = h_r^3$. However, for a fair comparison with our data-based choice of λ via GCV, the best choice for h_d is chosen by cross-validation via a grid search.

We investigate the regression model with a regular design i.e. $x_i = \frac{i}{n}, i = 1, \ldots, n$, normally distributed errors, sample size $n = 50$ and a signal to noise ratio (SNR) of 3. The signal-to-noise ratio is measured as $sd(f(x))/\sigma$, where $sd(f(x))$ is the estimated standard deviation of the regression function, $f(x_i)$ over the sample $i = 1, \ldots, n$, and σ is the true standard deviation of the noise in the data. The monotone regression functions that we

consider are (see Dette and Pilz [13])

$$m_1(x) = \frac{\exp(20(x - 1/2))}{1 + \exp(20(x - 1/2))}, \ m_2(x) = \frac{1}{2}(2x - 1)^3 + \frac{1}{2}, \ m_3(x) = x^2.$$

These functions correspond to, respectively, a function with a "continuous jump", a strictly increasing curve with a plateau, and a convex function. The different functions are displayed in Figures 1-3.

In Figures 1-3, we present some curves for the estimates of these three test functions. A Gaussian kernel has been used to compute the homeomorphic smoothing splines (using other kernels gives similar results). For the choice of the regularization parameter λ the GCV criterion (4.6) is used, and recall that we use cross-validation for the choice of h_d. As one can see in Figures 1-3, the homeomorphic smoothing spline based on the local linear estimator gives results similar to those obtained via the estimator of Dette, Neumeyer and Pilz [13]. However, our approach yields monotone estimator that are visually much smoother and very close to the true regression in all cases. Homeomorphic smoothing splines also seems to give very nice results even if the unconstrained estimator $\hat{f}_n$ is very oscillating as it is the case for the local linear estimator in Figure 2 and Figure 3.
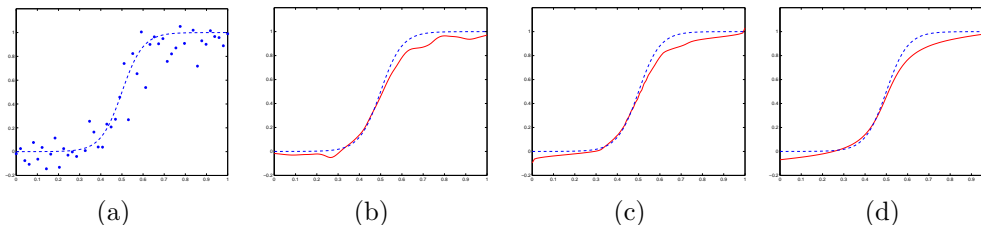

(a) (b) (c) (d)

Figure 1: *Signal m_1: the dotted line is the unknown regression function, (a) noisy data with $SNR = 3$, (b) Local Linear Unconstrained Estimator, (c) Dette et al.'s estimator, (d) Homeomorphic Smoothing Spline based on the Local Linear estimator.*
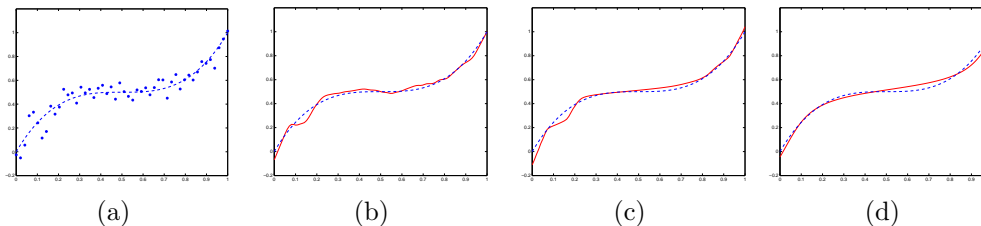

(a) (b) (c) (d)

Figure 2: *Signal m_2: the dotted line is the unknown regression function, (a) noisy data with $SNR = 3$, (b) Local Linear Unconstrained Estimator, (c) Dette et al.'s estimator, (d) Homeomorphic Smoothing Spline based on the Local Linear estimator.*

Table 1: Mean integrated squared error (MISE) over the 100 simulations for each method.

	Signal m_1	Signal m_2	Signal m_3
Homeomorphic smoothing spline	0.0032	0.00076	0.00089
Dette *et al.*'s estimator	0.0035	0.00098	0.0014

To compare these two monotone estimates, we have used 100 simulations runs for each regression function. The same unconstrained estimator (a local linear estimate with
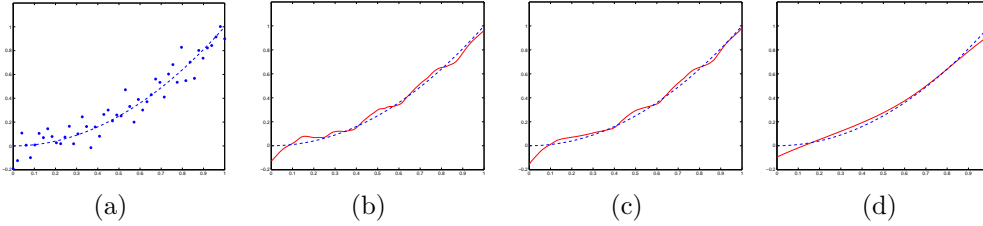
(a)   (b)   (c)   (d)

Figure 3: *Signal m_3: the dotted line is the unknown regression function, (a) noisy data with $SNR = 3$, (b) Local Linear Unconstrained Estimator, (c) Dette et al.'s estimator, (d) Homeomorphic Smoothing Spline based on the Local Linear estimator.*
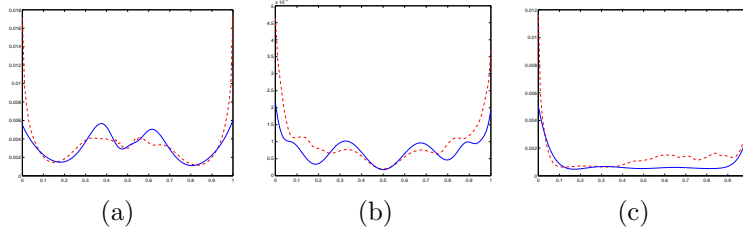

(a)   (b)   (c)

Figure 4: *Simulated mean squared error with $SNR = 3$ (computed over 100 simulations runs) on signal m_1, m_2 and m_3: Dette et al.'s estimator (dashed curves) and homeomorphic smoothing spline (solid curves) for the three regression functions.*

Epanechnikov kernel) is used. For the 100 simulations, we have calculated the pointwise mean squared error (MSE) for the two estimates $\hat{f}_n^c$ and $\hat{m}_n$, evaluated on an equidistant grid of size $2n$. Curves for the MSE of the three estimate are displayed in Figure 4. Again, these simulations clearly show that our approach compares similarly to the monotone estimator of Dette, Neumeyer & Pilz [13] for the signal m_1 and m_2, and outperforms Dette et al.'s estimator for the function m_3. Table 1 shows that it gives better results in terms of mean integrated squared error (MISE over $[0, 1]$) for the three test functions.

6.2   2D experiments and diffeormorphic matching

Let $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$ be two sets of n landmark in $\mathbb{R}^2$. The problem of landmark matching is to find a function $f : \mathbb{R}^2 \to \mathbb{R}^2$ such that $f(x_i) \approx y_i$ for all $i = 1, \ldots, n$ (see e.g Camion & Younes [10] and references therein). Let $\mathcal{H}_{K,2}$ be a RKHS of functions in $L^2(\mathbb{R}^2)$ with positive definite kernel K and denote by $\tilde{\mathcal{H}}_2$ the set of functions $f : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$f(x) = Ax + b + \begin{pmatrix} h_1(x) \\ h_2(x) \end{pmatrix}, \quad x \in \mathbb{R}^2,$$

where $h_1, h_2 \in \mathcal{H}_{K,2}$, A is 2×2 matrix and $b \in \mathbb{R}^2$. Landmark matching can be formulated as the problem of finding the minimizer

$$f_{n,\lambda} = \arg \min_{f \in \tilde{\mathcal{H}}_2} \frac{1}{n} \sum_{i=1}^n \|f(x_i) - y_i\|_{\mathbb{R}^2}^2 + \lambda(\|h_1\|_K^2 + \|h_2\|_K^2)$$

Under mild assumptions on the landmarks, the solution of this matching problem is unique and of the form: $\forall x \in \mathbb{R}, \; f_{n,\lambda}(x) = Ax + b + \begin{pmatrix} \sum_{i=1}^n \beta_{1,i} K(x, x_i) \\ \sum_{i=1}^n \beta_{2,i} K(x, x_i) \end{pmatrix}$, where A, b, β_1, β_2 are solutions of a simple linear system of equations (see e.g Camion & Younes

13

[10]). However, there are no constraints in this approach which guarantees that $f_{n,\lambda}$ is a one-to-one mapping of $\mathbb{R}^2$. Indeed, folding are possible for small values of λ as shown in Figure 5, where the mapping $f_{n,\lambda}$ is displayed via the deformation of an equally spaced grid of points in $\mathbb{R}^2$.

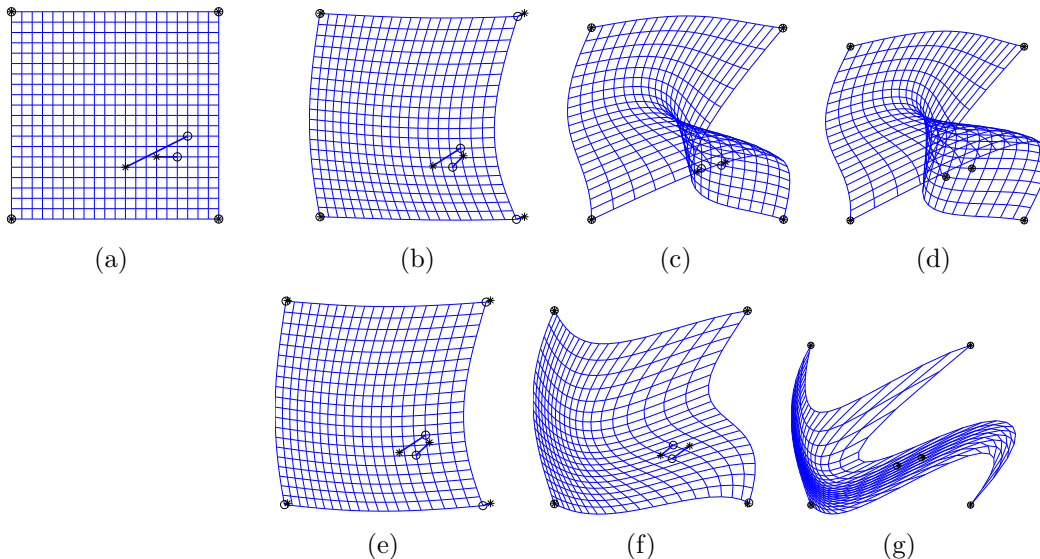Figure 5: *(a) $n = 6$ landmarks to be aligned $x_1, \ldots, x_n$ (circles) and $y_1, \ldots, y_n$ (stars). Four landmarks at the corner of the grid are already at the same location. The bold diagonal lines represents two landmarks be to aligned. Landmark matching with unconstrained spline smoothing: with a large λ (b), a moderate λ (c) and a small λ (d). Landmark matching with homeomorphic spline: with a large λ (e), a moderate λ (f) and a small λ (g).*

.

6.3 Homeomorphic spline for diffeomorphic matching

Let $\tilde{\mathcal{X}} = \{(v_t, t \in [0,1]) \text{ with } v_t \in \tilde{\mathcal{H}}_2 \text{ for all } t \in [0,1]\}$. Diffeomorphic matching of two sets of landmarks in $\mathbb{R}^2$ by homeomorphic spline is defined as the problem of finding a time-dependent vector field $v \in \tilde{\mathcal{X}}$, which minimizes the "energy"

$$\tilde{E}_\lambda(v) = \int_0^1 \frac{1}{n} \sum_{i=1}^n \|y_i - x_i - v_t(ty_i + (1-t)x_i)\|_{\mathbb{R}^2}^2 \, dt + \lambda \int_0^1 (\|h_{1,t}\|_K^2 + \|h_{2,t}\|_K^2) dt, \quad (6.1)$$

with $\lambda > 0$ a regularization parameter and $v_t(x) = A_t x + b_t + \begin{pmatrix} h_{1,t}(x) \\ h_{2,t}(x) \end{pmatrix}$, where for each $t \in [0,1]$, $h_{1,t}, h_{2,t} \in \mathcal{H}_{K,2}$, A_t is 2×2 matrix and $b_t \in \mathbb{R}^2$. Then, by taking $v^{n,\lambda} = \arg\min_{v \in \tilde{\mathcal{X}}} \tilde{E}_\lambda(v)$, a diffeomorphic mapping between these two sets of landmarks is obtained by computing $f_{n,\lambda}^c(x) = \phi_1^{v^{n,\lambda}}(x) = x + \int_0^1 v_t^{n,\lambda}(\phi_t^{v^{n,\lambda}}(x)) dt$. Under mild assumptions, and arguing as in the proof of Proposition 4.1, the optimization problem (6.1) has a unique solution $v^{n,\lambda} \in \tilde{\mathcal{X}}$ such that at each time $t \in [0,1]$, $v_t^{n,\lambda}$ is the solution of the following standard unconstrained smoothing problem: find $v_t \in \tilde{\mathcal{H}}_2$ which minimizes $\tilde{E}_\lambda^t(v_t) = \frac{1}{n} \sum_{i=1}^n \|y_i - x_i - v_t(ty_i + (1-t)x_i)\|_{\mathbb{R}^2}^2 + \lambda(\|h_{1,t}\|_K^2 + \|h_{2,t}\|_K^2)$, where $v_t(x) = A_t x + b_t + \begin{pmatrix} h_{1,t}(x) \\ h_{2,t}(x) \end{pmatrix}$. Hence, the computation of the diffeomorphic mapping $f_{n,\lambda}^c$ is obtained using unconstrained spline smoothing and by running an ODE. Numerically, we use an Euler scheme similar to the one proposed in Section 4.3.

Remark that the formulation (6.1) is somewhat similar to the geodesic smoothing spline problem proposed by Camion & Younes [10] in 2D setting. To compute a smooth diffeomorphism to align two sets of landmarks $(x_i, y_i), i = 1, \ldots, n$ in $\mathbb{R}^2 \times \mathbb{R}^2$, Camion & Younes [10] suggest to minimize the following energy

$$\sum_{i=1}^{n} \|\frac{dq_i(t)}{dt} - v_t(q_i(t)\|_{\mathbb{R}^2}^2 + \lambda \int_0^1 (\|h_{1,t}\|_K^2 + \|h_{2,t}\|_K^2)dt \qquad (6.2)$$

over all time-dependent vector fields and all landmark trajectories $q_i(t), i = 1, \ldots, n$ with initial conditions $q_i(0) = x_i$ and $q_i(1) = y_i$. This leads to an optimization problem which can be solved by a gradient-descent algorithm. In our formulation (6.1), the landmarks trajectories are fixed, and correspond to linear paths $q_i(t) = ty_i + (1 - t)x_i$ between x_i and y_i. This makes the optimization problem (6.2) easier to solve.

An example of diffeomorphic mapping is shown in Figure 5, and one can see that even for small values of λ the mapping remains one-to-one contrary to unconstrained spline smoothing. An example of landmark-based image warping is also displayed in Figure 6 which illustrates the advantages of homeospline over unconstrained spline smoothing which may lead to unrealistic matching.

(a)	(b)	(c)	(d)

Figure 6: *Image warping based on landmark alignment. Figures (a) and (b) are two images with manually annotated landmarks representing facial structures. Figure (c) is the warping of image (b) onto image (a) using landmark alignment with unconstrained spline smoothing. Figure (d) is the warping of image (b) onto image (a) using landmark alignment with homeomorphic spline. Images are taken from the IMM Face Database [37].*

7   Conclusion and Future works

Homeomorphic splines allows one to compute with a low computational cost, monotone regressors in a 1D setting. In the presence of noisy data, this leads to an estimator which performs at least as well as existing ones. Moreover such an estimator has an optimal rate of convergence over a large class of functional spaces. Homeomorphic splines can also be extended in a 2D setting for landmarks and image warping. In future work, we plan to investigate applications for 2D regression under monotonicity constraints, and also to study the asymptotic normality of our estimator.

Appendix

Proof of Theorem 3.1: the key point to derive this result is the use of the following lemma which follows immediately from condition (3.4) and the assumption $M = 2$, $\psi_1(x) = 1$, $\psi_2(x) = x$ made in Section 2.

Lemma 7.1 *Suppose that the assumptions of Theorem 3.1 are satisfied. Then, for any $g \in \tilde{\mathcal{H}}$, there exists a constant C_2 (not depending on g) such that for all $x, y \in \mathbb{R}$, $|g(x) - g(y)| \leq C_2\|g\||x - y|$, and $|g(x)| \leq C_2\|g\|(1 + |x|)$.*

Existence and uniqueness of the solution of the ODE : It follows from the Picard-Lindelof Theorem (see Theorem 19 and 20 in Younes [44] using the same notations).

Invertibility of the solution : for $v \in \mathcal{X}^1$, $x \in \Omega$ and any $t, s \in [0, 1]$, we denote by $\phi_{st}^v(x)$ the value at time t of the unique solution of the equation $\frac{d\phi_r}{dr} = v_r(\phi_r)$ which is equal to x at time s. The solution $\phi_{st}^v(x)$ is equal to $\tilde{x} = \phi_{sr}^v(x)$ at time $r \in [0, 1]$, and thus also equal to $\phi_{rt}^v(\tilde{x}) = \phi_{st}^v(x)$ which implies that $\phi_{st}^v(x) = \phi_{rt}^v \circ \phi_{sr}^v(x)$ and thus $\phi_{st}^v \circ \phi_{ts}^v(x) = x$ which proves that for all $t \in [0, 1]$ $\phi_t = \phi_{0t}$ is invertible, and that its inverse is given by $\phi_t^{-1} = \phi_{t0}$.

Continuity of ϕ_t and its inverse: ϕ_t is a homeomorphism using Gronwall's lemma and the arguments in Younes [44]. This ends the proof of Theorem 3.1. $\square$

Proof of Theorem 4.1: let $t \in [0, 1]$ and define the interval $I_t = [tf(0), tf(1) + (1-t)] = [a_t, b_t]$. Given our assumptions on f the function $\phi_t(x) = tf(x) + (1-t)x$ is a continuously differentiable homeomorphism from $[0, 1]$ to I_t such that its inverse is continuous and differentiable with $\frac{d\phi_t^{-1}}{dy}(y) = \frac{1}{tf'(\phi_t^{-1}(y)) + (1-t)}$ for all $y \in I_t$, which shows that $\frac{d\phi_t^{-1}}{dy}$ is continuous since we have assumed that $f'(x) > 0$ for all $x \in [0, 1]$. Hence, ϕ_t is a diffeomorphism from $[0, 1]$ to I_t. Moreover, the above formula for $\frac{d\phi_t^{-1}}{dy}$ and the fact that $f \in H^m([0, 1])$ implies that ϕ_t^{-1} belongs to $H^m(I_t)$ by application of the chain rule of differentiation. Let $\tilde{f}_t$ the function defined on I_t such that $\tilde{f}_t(y) = f(\phi_t^{-1}(y)) - \phi_t^{-1}(y)$ for all $y \in I_t$. Given that ϕ_t is a diffeomorphism from $[0, 1]$ to I_t and that $\phi_t^{-1} \in H^m(I_t)$, we can again apply the chain rule of differentiation to show that $\tilde{f}_t$ belongs to $H^m(I_t)$. Now, define the sub-space H_0 of functions in $H^m(\mathbb{R})$ which coincide with $\tilde{f}_t$ on $I_t = [a_t, b_t]$. First, we shall prove that H_0 is not empty. Indeed, choose $c_t < a_t$ and $d_t > b_t$ and define the Hermite polynomials P_0 and P_1 of degree $2m + 1$ such that for all $0 \leq k \leq m - 1$, $P_0^{(k)}(a_t) = \tilde{f}_t^{(k)}(a_t)$ and $P_0^{(k)}(c_t) = 0$, and $P_1^{(k)}(b_t) = \tilde{f}_t^{(k)}(b_t)$ and $P_1^{(k)}(d_t) = 0$. Then, define the function v_t^m on $\mathbb{R}$ such that

$$v_t^m(y) = \begin{cases} 0 & \text{for } y \in]-\infty, c_t[, \\ P_0(y) & \text{for } y \in [c_t, a_t[, \\ \tilde{f}_t(y) & \text{for } y \in [a_t, b_t], \\ P_1(y) & \text{for } y \in]b_t, d_t], \\ 0 & \text{for } y \in]d_t, +\infty[, \end{cases}$$

By construction of P_0 and P_1 and the fact that $\tilde{f}_t \in H^m(I_t)$, we have that v_t^m belongs to $H^m(\mathbb{R})$, and thus H_0 is not empty. Since the space H_0 is closed and convex, it contains a unique element of minimum norm that we denote by v_t^f which satisfies equation (4.1) by construction of H_0, which completes the proof Theorem 4.1. $\square$

Proof of Proposition 4.1 : let $t \in [0, 1]$. Given our assumptions on the x_i's and the y_i's we can define the function $v_t^{n,\lambda} \in \tilde{\mathcal{H}}$ as the smoothing spline which minimizes the energy

16

$E_\lambda^t(v_t)$ as defined in equation (4.3). Then, by definition of $v_t^{n,\lambda}$ we have that for any $v_t \in \tilde{\mathcal{H}}$, $E_\lambda^t(v_t^n) \le E_\lambda^t(v_t)$ which implies that for any $v \in \mathcal{X}$ $E_\lambda(v^{n,\lambda}) = \int_0^1 E_\lambda^t(v_t^n)dt \le E_\lambda(v)$, which proves that $v^{n,\lambda}$ is a minimum of $E_\lambda(v)$. Then, the uniqueness of $v^{n,\lambda}$ follows from the strict convexity of E_λ.

Now, for $i = 1, \ldots, n$ and $t \in [0, 1]$, let $\hat{X}_i^t = ty_i + (1 - t)x_i$ and $\hat{Y}_i = y_i - x_i$. Then, the time-dependent vector field $v_t^{n,\lambda}$ is such that for any $x \in \mathbb{R}$ (see Wahba [42]) $v_t^{n,\lambda}(x) = \alpha_1^t + \alpha_2^t x + \sum_{i=1}^n \beta_i^t K(x, \hat{X}_i^t)$, where the coefficients $\boldsymbol{\alpha}_t = (\alpha_1^t, \alpha_2^t)'$ and $\boldsymbol{\beta}_t = (\beta_1^t, \ldots, \beta_n^t)'$ are given by: $\boldsymbol{\beta}_t = \hat{\Sigma}_{\lambda,t}^{-1}(I_n - P_t)\mathbf{Y}$ and $\boldsymbol{\alpha}_t = (T_t'\hat{\Sigma}_{\lambda,t}^{-1}T_t)^{-1}T_t'\hat{\Sigma}_{\lambda,t}^{-1}\mathbf{Y}$, with $\mathbf{Y} = (\hat{Y}_1, \ldots, \hat{Y}_n)'$, $\hat{\Sigma}_{\lambda,t} = \hat{\Sigma}_t + n\lambda I_n$ and $\hat{\Sigma}_t$ is the $n \times n$ matrix with elements $\hat{\Sigma}_t[i, j] = K(\hat{X}_i^t, \hat{X}_j^t)$, T_t is the $n \times 2$ matrix with elements $T_t[i, 1] = 1$ and $T_t[i, 2] = \hat{X}_i^t$, and P_t is the $n \times n$ matrix given by $P_t = T_t(T_t'\hat{\Sigma}_{\lambda,t}^{-1}T_t)^{-1}T_t'\hat{\Sigma}_{\lambda,t}^{-1}$. Then by the continuity of K and the continuity of matrix inversion, it follows that the coefficients $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$ are continuous functions of t on $[0, 1]$. Hence, $t \mapsto \|v_t^{n,\lambda}\|$ is a continuous function on $[0, 1]$ which implies that $\int_0^1 \|v_t^{n,\lambda}\|^2 dt < +\infty$. Hence $v^{n,\lambda} \in \mathcal{X}^2$ which completes the proof of Proposition 4.1 using Theorem 3.1. $\square$

Proof of Theorem 5.1: in the proof, C will denote a constant whose value may change from line to line. For $i = 1, \ldots, n$ and $t \in [0, 1]$, let $\hat{X}_i^t = t\hat{f}^n(x_i) + (1 - t)x_i$, $X_i^t = tf(x_i) + (1 - t)x_i$ and $\hat{Y}_i = \hat{f}^n(x_i) - x_i$. One can remark that the smoothing spline $v_t^{n,\lambda}$ evaluated at the "design points" $\hat{X}_1^t, \ldots, \hat{X}_n^t$ is a linear function of the observations $\hat{Y}_1, \ldots, \hat{Y}_n$ and can therefore be written as $\mathbf{v}_t^{n,\lambda} = \left(v_t^{n,\lambda}(\hat{X}_1^t), \ldots, v_t^{n,\lambda}(\hat{X}_n^t)\right)' = A_{\lambda,t}\mathbf{Y}$, where $A_{\lambda,t} = P_t + \hat{\Sigma}_t\hat{\Sigma}_{\lambda,t}^{-1}(I_n - P_t)$. Then, under the assumptions of Theorem 5.1, the following lemma holds (the proof follows using arguments in Craven & Wahba [11]):

Lemma 7.2 *For almost all $t \in [0, 1]$ and any $\mathbf{x} \in \mathbb{R}^n$, we have $\|A_{\lambda,t}\mathbf{x}\|_2 \le 2\|\mathbf{x}\|_2$ a.s.*

The next step shows that $v_t^{n,\lambda}$ is in $\mathcal{X}^2$ with an asymptotically probability equals to 1.

Lemma 7.3 *There exists $C_5 > 0$ such that $\lim_{n \to +\infty} \mathbb{P}\left(\int_0^1 \|v_t^{n,\lambda}\|^2 dt \le C_5\right) = 1$.*

Proof: recall that $v^{n,\lambda}$ is the minimizer of the following energy:

$E_{\lambda_n}(v) = \int_0^1 \frac{1}{n}\sum_{i=1}^n \left(\hat{f}^n(x_i) - x_i - v_t(t\hat{f}^n(x_i) + (1 - t)x_i)\right)^2 dt + \lambda_n \int_0^1 \|h_t\|_K^2 dt$, where $v_t(x) = a_1^t + a_2^t x + h_t(x)$. We will show that $v^{n,\lambda}$ converges in probability to v^f for the norm $\|v\|_{\mathcal{X}^2} = \int_0^1 \|v_t\|^2 dt$. Let $E(v) = \int_0^1 \int_0^1 (f(x) - x - v_t(tf(x) + (1 - t)x))^2 w(x)dxdt$, $\mathcal{A}$ be a compact set in $\mathcal{X}^2$ and take $v \in \mathcal{A}$: $(E_{\lambda_n}(v) - E(v))^2 \le 2E_{1,n}^2(v) + 2E_{2,n}^2(v)$, where $E_{2,n}(v) = \lambda_n \int_0^1 \|h_t\|_K^2 dt$ and

$$
\begin{aligned}
E_{1,n}(v) &= \int_0^1 \int_0^1 (f(x) - x - v_t(tf(x) + (1 - t)x))^2 (dw_n(x) - w(x)dx)dt + \\
&\quad \int_0^1 \int_0^1 \left[\left(\hat{f}_n(x) - x - v_t(t\hat{f}_n(x) + (1 - t)x\right)^2 - \\
&\quad (f(x) - x - v_t(tf(x) + (1 - t)x))^2\right] dw_n(x)dt,
\end{aligned}
$$

with $w_n(x) = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}(x)$, since $\lambda_n \to 0$, we have that $\sup_{v \in \mathcal{A}} E_{2,n}^2(v) \to 0$ as $n \to +\infty$. Then, remark that $E_{1,n}^2(v) \le 2I_{1,n}^2(v) + 2I_{2,n}^2(v)$, where $I_{1,n}(v) = \int_0^1 \int_0^1 (f(x) - x - v_t(tf(x) + (1 - t)x))^2 (dw_n(x) - w(x)dx)dt$, and $I_{2,n}(v) = \int_0^1 \int_0^1 \left[\left(\hat{f}_n(x) - x\right)$

$$-v_t(t\hat{f}_n(x) + (1-t)x)\Big)^2 - (f(x) - x - v_t(tf(x) + (1-t)x))^2\Big]\, dw_n(x)dt.$$ Let $g_t^v(x) = (f(x) - x - v_t(tf(x) + (1-t)x))^2$, then $I_{1,n}(v) = \int_0^1 \int_0^1 g_t^v(x)dt(dw_n(x) - w(x)dx)$ by Fubini theorem. Lemma 7.1, implies that $x \mapsto \int_0^1 g_t^v(x)dt$ is bounded on $[0,1]$ by $C_v = \int_0^1 (1 + \|f\|_\infty + C_2\|v_t\|(2 + \|f\|_\infty))^2\, dt$. Then, from the compactness of $\mathcal{A}$ it follows that there exists a constant C such that for all $v \in \mathcal{A}$ $I_{1,n}(v) \le C \int_0^1 (dw_n(x) - w(x)dx)$. Hence, by definition of $w(x)$ and $w_n(x)$ the inequality above finally implies that

$$\sup_{v \in \mathcal{A}} I_{1,n}^2(v) \to 0 \text{ as } n \to +\infty. \tag{7.1}$$

Now, using Cauchy-Schwarz inequality we have that $I_{2,n}^2 \le I_{3,n}(v)I_{4,n}(v)$, where

$$I_{3,n}(v) = \int_0^1 \int_0^1 \left(\hat{f}_n(x) - f(x) - v_t(t\hat{f}_n(x) + (1-t)x) + v_t(tf(x) + (1-t)x)\right)^2 dw_n(x)dt,$$

and $I_{4,n}(v) = \int_0^1 \int_0^1 \left(\hat{f}_n(x) + f(x) - 2x - v_t(t\hat{f}_n(x) + (1-t)x) - v_t(tf(x) + (1-t)x)\right)^2 dw_n(x)dt.$ Note that using Lemma 7.1, it follows that

$$I_{4,n}(v) \le 2R_n(\hat{f}_n, f) + 4\int_0^1 (\|f\|_\infty + 2 + 2C_2\|v_t\|(2 + \|f\|_\infty))^2\, dt + 4C_2^2 R_n(\hat{f}_n, f) \int_0^1 \|v_t\|^2 dt.$$

Then, one has that

$$I_{3,n}(v) \le 2\int_0^1 \int_0^1 \left(\hat{f}_n(x) - f(x)\right)^2 dw_n(x)dt + 2C_2^2 \int_0^1 \int_0^1 \|v_t\|^2 \left(\hat{f}_n(x) - f(x)\right)^2 dw_n(x)dt,$$

and thus $\sup_{v \in \mathcal{A}} I_{3,n}(v) \le 2R_n(\hat{f}_n, f) + 2C_2^2 \sup_{v \in \mathcal{A}} \int_0^1 \|v_t\|^2 dt R_n(\hat{f}_n, f)$. By assumption, $R_n(\hat{f}_n, f) \to 0$ in probability, and thus $\sup_{v \in \mathcal{A}} I_{3,n}(v) \to 0$ in probability as $n \to +\infty$. By combining the above equation with the bound for $I_{4,n}(v)$, we finally obtain that in probability $\sup_{v \in \mathcal{A}} I_{2,n}^2(v) \to 0$ as $n \to +\infty$. So finally, we obtain from (7.1) that $\sup_{v \in \mathcal{A}} E_{1,n}^2(v) \to 0$ in probability as $n \to +\infty$, which implies together with equation (7) that $\sup_{v \in \mathcal{A}} (E_{\lambda_n}(v) - E(v))^2 \to 0$ in probability as $n \to +\infty$. Now, remark that since $E_{\lambda_n}(v)$ and $E(v)$ are positive and strictly convex functionals, they have a unique minimum over the set of time-dependent vector fields $\mathcal{X}$. Moreover, by definition of $v^f \in \mathcal{X}^2$ one has that for any $t \in [0,1]$, $v_t^f(tf(x) + (1-t)x) = f(x) - x$ which implies that v^f is the minimum of $E(v)$. Let $\epsilon > 0$ and define $B(v^f, \epsilon) = \{v \in \mathcal{X}^2 ; \|v - v^f\|_{\mathcal{X}^2} \le \epsilon\}$, and let $dB(v^f, \epsilon) = \{v \in \mathcal{X}^2 ; \|v - v^f\|_{\mathcal{X}^2} = \epsilon\}$ be the frontier of $B(v^f, \epsilon)$. Since v^f is the minimum of $E(v)$, there exits $\delta > 0$ such that for any $v \in dB(v^f, \epsilon)$ $E(v^f) < E(v) + \delta$. Obviously, $B(v^f, \epsilon)$ is a compact subset of $\mathcal{X}^2$, and thus $\sup_{v \in B(v^f, \epsilon)} |E_{\lambda_n}(v) - E(v)|$ converges to zero in probability. This implies that for any $\alpha > 0$, there exists $n_1 \in \mathbb{N}$ such that for any $n \ge n_1$, $\mathbb{P}\left(v \in dB(v^f, \epsilon) ; E_{\lambda_n}(v) - E(v) > -\frac{\delta}{3}\right) \ge 1 - \frac{\alpha}{2}$. Similarly, there exists n_2 such that for any $n \ge n_2$, $\mathbb{P}\left(E(v^f) - E_{\lambda_n}(v^f) > -\frac{\delta}{3}\right) \ge 1 - \frac{\alpha}{2}$. Hence, we have that for any $n \ge \max(n_1, n_2)$, $\mathbb{P}\left(v \in dB(v^f, \epsilon) ; E_{\lambda_n}(v) > E_{\lambda_n}(v^f) + \frac{\delta}{3}\right) \ge 1 - \alpha$. This implies that except on a set of probability less than α, E_{λ_n} has a local minimum in the interior of $B(v^f, \epsilon)$ which is thus the global minimum $v^{n,\lambda}$ since E_{λ_n} is strictly convex. Hence, we finally have that for any $\epsilon > 0$ then with probability tending to one $v^{n,\lambda}$ belongs to $B(v^f, \epsilon)$ which implies that $v^{n,\lambda}$ converges in probability to v^f for the norm $\|\cdot\|_{\mathcal{X}^2}$. This proves that there exists a constant C_5 (not depending on n) such that as $n \to +\infty$,

$\mathbb{P}(\|v^{n,\lambda}\|_{\mathcal{X}^2} \leq C_5) \to 1$, which completes the proof of Lemma 7.3. $\square$

Now, since $v^{n,\lambda} \in \mathcal{X}^2$, Theorem 3.1 implies that we can define $\hat{f}_{n,\lambda}$ and $\phi_t^{n,\lambda}$ as the solutions respectively at time $t=1$ and time $t \in [0,1]$ of the ODE $\frac{d\phi_t}{dt} = v_t^{n,\lambda}(\phi_t)$. We shall now control the empirical error $R_n(\hat{f}_n^c, f) = \frac{1}{n}\sum_{i=1}^{n}(\hat{f}_n^c(x_i) - f(x_i))^2$. First, $\sum_{i=1}^{n}(\hat{f}_n^c(x_i) - f(x_i))^2 = \sum_{i=1}^{n}(\phi_1^{n,\lambda}(x_i) - \phi_1^f(x_i))^2 = \sum_{i=1}^{n}\left(\int_0^1 \left(v_t^{n,\lambda}(\phi_t^{n,\lambda}(x_i)) - v_t^f(\phi_t^f(x_i))\right)dt\right)^2$. and note that for any $t \in [0,1]$

$$\sum_{i=1}^{n}(\phi_t^{n,\lambda}(x_i) - \phi_t^f(x_i))^2 = \sum_{i=1}^{n}\left(\int_0^t \left(v_s^{n,\lambda}(\phi_s^{n,\lambda}(x_i)) - v_s^f(\phi_s^f(x_i))\right)ds\right)^2,$$

which implies that (using Cauchy-Schwarz inequality and the fact that $t \leq 1$)

$$\sum_{i=1}^{n}(\phi_t^{n,\lambda}(x_i) - \phi_t^f(x_i))^2 \leq 2\sum_{i=1}^{n}\int_0^t \left(v_s^{n,\lambda}(\hat{X}_i^s) - v_s^f(\phi_s^f(x_i))\right)^2 ds$$

$$+ \int_0^t 4C_2^2\|v_s^{n,\lambda}\|^2\sum_{i=1}^{n}\left(\hat{X}_i^s - \phi_s^f(x_i)\right)^2 ds + \int_0^t 4C_2^2\|v_s^{n,\lambda}\|^2\sum_{i=1}^{n}\left(\phi_s^{n,\lambda}(x_i) - \phi_s^f(x_i)\right)^2 ds.$$

To bound this sum, we shall use the following Lemma whose proof can be found in Younes [44]

Lemma 7.4 *Consider three continuous and positive functions* c_s, γ_s *and* u_s *defined on* $[0,1]$ *and such that* $u_t \leq c_t + \int_0^t \gamma_s u_s ds$, *then* $u_t \leq c_t + \int_0^t c_s \gamma_s e^{\int_s^t \gamma_r dr} ds$.

Then, if we apply Lemma 7.4 by letting $u_t = \sum_{i=1}^{n}(\phi_t^{n,\lambda}(x_i) - \phi_t^f(x_i))^2$, $\gamma_t = 4C_2^2\|v_t^{n,\lambda}\|^2$, $c_t = 2\int_0^t \sum_{i=1}^{n}\left(v_s^{n,\lambda}(\hat{X}_i^s) - v_s^f(\phi_s^f(x_i))\right)^2 ds + \int_0^t 4C_2^2\|v_s^{n,\lambda}\|^2\sum_{i=1}^{n}\left(\hat{X}_i^s - \phi_s^f(x_i)\right)^2 ds$, we obtain that $u_t \leq c_t + \int_0^t c_s\gamma_s e^{\int_s^t \gamma_r dr}ds$. Now recall that by definition of v_s^f and ϕ_s^f, we have $\phi_s^f(x_i) = X_i^s$ and $v_s^f(\phi_s^f(x_i)) = f(x_i) - x_i =_{def} \tilde{f}(x_i)$.

Hence, $\sum_{i=1}^{n}\left(v_s^{n,\lambda}(\hat{X}_i^s) - v_s^f(\phi_s^f(x_i))\right)^2 = \|A_{\lambda,s}\mathbf{Y} - \tilde{\mathbf{f}}\|_2^2$, since by definition of $A_{\lambda,s}$ one has $\left(v_s^{n,\lambda}(\hat{X}_1^s), \ldots, v_s^{n,\lambda}(\hat{X}_n^s)\right)' = A_{\lambda,s}\mathbf{Y}$, and where $\tilde{\mathbf{f}} = \left(\tilde{f}(x_1), \ldots, \tilde{f}(x_n)\right)'$. Similarly, we have that $\sum_{i=1}^{n}\left(\hat{X}_i^s - \phi_s^f(x_i)\right)^2 = s^2\|\hat{\mathbf{f}}_n - \mathbf{f}\|_2^2 \leq \|\hat{\mathbf{f}}_n - \mathbf{f}\|_2^2$, where $\hat{\mathbf{f}}_n = \left(\hat{f}_n(x_1), \ldots, \hat{f}_n(x_n)\right)'$ and $\mathbf{f} = (f(x_1), \ldots, f(x_n))'$. Hence,

$$c_t \leq 2\int_0^t \|A_{\lambda,s}\mathbf{Y} - \tilde{\mathbf{f}}\|_2^2 ds + \int_0^t 4C_2^2\|v_s^{n,\lambda}\|^2 ds\|\hat{\mathbf{f}}_n - \mathbf{f}\|_2^2. \tag{7.2}$$

Now, remark that $\|A_{\lambda,s}\mathbf{Y} - \tilde{\mathbf{f}}\|_2^2 \leq \|(I_n - A_{\lambda,s})\tilde{\mathbf{f}}\|_2^2 + \|A_{\lambda,s}(\mathbf{Y} - \tilde{\mathbf{f}})\|_2^2$, and observe that by Lemma 7.2

$$\|A_{\lambda,s}(\mathbf{Y} - \tilde{\mathbf{f}})\|_2^2 \leq 2\|\hat{\mathbf{f}}_n - \mathbf{f}\|_2^2, \tag{7.3}$$

and let $\tilde{v}_s^{n,\lambda}$ be the solution of the following smoothing problem: find $v \in \tilde{\mathcal{H}}$ which minimizes $\frac{1}{n}\sum_{i=1}^{n}\left(\tilde{f}(x_i) - v(\hat{X}_i^s)\right)^2 + \lambda\|h\|_K^2$, where $v_t(x) = a_1^t + a_2^t x + h_t(x)$. Then, by definition of $\tilde{v}_s^{n,\lambda}$ we have that

$$\frac{1}{n}\|(I_n - A_{\lambda,s})\tilde{\mathbf{f}}\|_2^2 + \lambda\|\tilde{h}_s^{n,\lambda}\|_K^2 \leq \frac{1}{n}2C_2^2\|v_s^f\|^2\sum_{i=1}^{n}\left(X_i^t - \hat{X}_i^t\right)^2 + \lambda\|h_s^f\|_K^2$$

19

Finally, by combing equations (7.3) and the above inequality, we obtain that there exists a constant C such that

$$\|A_{\lambda,s}\mathbf{Y} - \tilde{\mathbf{f}}\|_2^2 \leq C\left(\|\hat{\mathbf{f}}_n - \mathbf{f}\|_2^2 + n\lambda\right) \text{ with } C = \max(2C_2^2\|v_s^f\|^2, \|h_s^f\|_K^2). \qquad (7.4)$$

Now using Lemma 7.3, and combining the above relation with equations (7.2) and (7.4), we finally obtain in probability $c_t \leq b(n)$. where $b(n) = C\left(\|\hat{\mathbf{f}}_n - \mathbf{f}\|_2^2 + n\lambda\right)$ for some constant $C > 0$. Similarly, by Lemma 7.3 we have that there exists a constant C_6 such that (in probability) for any $s, t \in [0,1]$ $\int_s^t \gamma_r dr \leq C_6$. Then, by combining the previous inequalities, we derive from Lemma (7.4) that $u_t \leq b(n)(1 + C_6 e^{C_6})$. Now, since $R_n(\hat{f}_n^c, f) = \frac{1}{n}u_1$ and $R_n(\hat{f}_n, f) = \frac{1}{n}\|\hat{\mathbf{f}}_n - \mathbf{f}\|_2^2$, we finally obtain that there exists a constant Λ_1 such that with probability tending to one as $n \to +\infty$ $R_n(\hat{f}_n^c, f) \leq \Lambda_1\left(R_n(\hat{f}_n, f) + \lambda_n\right)$, which completes the proof of Theorem 5.1. $\square$

References

[1] Andersson, L. E. and Elfving, T. (1991). Interpolation and Approximation by Monotone Cubic Splines, *J. Approx. Theory* (66), 302-333.

[2] Apprato, D. and Gout, C. (2000). A result about scale transformation families in approximation: application to surface fitting from rapidly varying data, Numerical Algorithms 23 (2,3), 263-279.

[3] Antoniadis, A., Bigot, J. and Gijbels, I. (2007). Penalized wavelet monotone regression, *Statistics and Probability Letters*, (77), 1608-1621.

[4] Aronszajn N. (1950). Theory of reproducing kernels. Transactions of the American Mathematical Society (68): 337-404.

[5] Atteia, M. (1992). Hilbertian Kernels and splines functions . Studies in computational mathematics 4, North-Holland.

[6] Atteia, M. and Gaches, J. (1999). Approximation hilbertienne: Splines. Ondelettes, Fractales. Presses Universitaires de Grenoble, Grenoble.

[7] Beatson R.K. and Ziegler Z. (1985). Monotonicity preserving surface interpolation, SIAM J. Numer. Anal., 22, 401-411.

[8] Berlinet, A. and Thomas-Agnan, C (2004). Reproducing Kernel Hilbert Spaces in Probability and Statistics. *Kluwer Academic Publishers, Boston.*

[9] Bigot, J., Gadat, S., and Loubes J.M. (2009) Statistical M-Estimation and Consistency in large deformable models for Image Warping, Journal of Mathematical Imaging and Vision, 34 (3), pp. 270-290.

[10] Camion, V. and Younes, L. (2001). Geodesic Interpolating Splines, *Energy Minimization Methods for Computer Vision and Pattern Recognition (EMMCVPR), New York.* 513-527.

[11] Craven, P. and Wahba, G. (1979), Smoothing noisy data with spline function, *Numerische Mathematik*, vol. 31, pp. 377–403.

[12] de Boor, C. and Schumaker L. (1976), On calculating with B splines II, Integration, *Numerische Methoden der Approximations Theorie*, **3**, 123–146.

[13] Dette, H. , Neumeyer, N. and Pilz, K.F. (2006), A simple nonparametric estimator of a strictly monotone regression function, *Bernoulli*, **12**, no. 3, 469–490.

[14] Dette, H. and Pilz, K.F. (2006). A comparative study of monotone nonparametric kernel estimates, *Journal of Statistical Computation and Simulation*, **76**, no. 1, 41–56.

[15] Duchon, J. (1977). Splines minimizing rotation-invariant semi-norms in Sobolev spaces, *Lecture Notes in Mathematics* , **571**, 85–100.

[16] Elfving T. and Andersson (1989), L. E.. An Algorithm for Computing Constrained Smoothing Spline Functions, *Numer. Math.*, 52, 583–595.

[17] Fan J. and Gijbels I. (1996). Local polynomial modelling and its applications. *Chapman and Hall, London.*

[18] Gijbels, I. (2005). Monotone regression, to appear in *Encyclopedia of Statistical Sciences, Second Edition.* Editors S. Kotz, N.L. Johnson, C.B. Read, N. Balakrishnan, and B. Vidakovic. Wiley, New York.

[19] Glaunès, J. (2005). Modèles déformables en appariement de formes. Transport par difféomorphismes de points, de mesures et de courants pour la comparaison de formes et l'anatomie numérique, *Phd Thesis (in french), Université Paris 13.*

[20] Girard (1995). The Fast Monte-Carlo Cross-Validation and C_L Procedures: Comments, New Results and Application to Image Recovery Problems, *Computational Statistics*, **10**, 205–231.

[21] Hall, P. and Huang L.-S. (2001). Nonparametric kernel regression subject to monotonicity constraints, *The Annals of Statistics*, **29**, 624-647.

[22] He, X. and Shi, P. (1998) Monotone B-Spline Smoothing. *Journal of the American Statistical Association*, **14**, 315-337

[23] Irvine, L.D., Marin, S.P. and Smith, P.W. (1986). Constrained interpolation and smoothing, *Constructive Approximation* 2(2): 129?151.

[24] Kelly, C. and Rice, J. (1990). Monotone smoothing with application to dose response curves and the assessment of synergism, *Biometrics*, **46**, 1071–1085.

[25] Kozlov, M.K. and Tarasov, S.P. and Khachiyan, L.G. (1979). Polynomial solvability of convex quadratic programming, *Sov. Math., Dokl.* **20** , 1108-1111.

[26] Lopez de Silanes, M.C. and Apprato, D. (1988). Approximation error estimates on a bounded domain in $\mathbb{R}^n$ interpolating and smoothing discrete D^m-splines. *Numerische Mathematik*, **53**, (3) 367–376.

[27] Mammen, E. (1991). Estimating a smooth monotone regression function, *The Annals of Statistics*, **19**, 724–740.

[28] Mammen, E. , Marron, J.S. , Turlach, B.A. and Wand, M.P. (2001). A general projection framework for constrained smoothing, *Statist. Sci.*, **16**, 232–248.

[29] Mammen, E. and Thomas-Agnan, C. (1999). Smoothing splines and shape restrictions, *Scand. J. Statist.* , **26**, 239–252.

[30] Miller, M.I., Trouvé, A. and Younes, L. (2002). Group actions, homeomorphisms, and matching: A general framework, *International Journal of Computer Vision*, 41(1/2):61–84.

[31] Nussbaum, M. (1985). Spline smoothing in regression models and asymptotic efficiency in L2, *The Annals of Statistics*, **13**, 984-992.

[32] Pardalos, P.M., and Vavasis, S.A. (1991). Quadratic programming with one negative eigenvalue is NP-hard . *Journal of Global Optimization*, **1**, (1)15–22.

[33] Quak, E. and Schumaker, L. (1991). Least squares fitting by linear splines on data dependent triangulations. *in* Curves and Surfaces, *Academic Press, N. Y.*, 387–390.

[34] Renardy, M. and Rogers, R. (2004). An Introduction to Partial Differential Equations, *Springer, Berlin.*

[35] Schwetlick, H. and Kunert, V. (1993). Spline smoothing under constraints on derivatives, *Bit*, **33**, 512?528.

[36] Speckman, P. (1985). Spline smoothing and optimal rates of convergence in nonparametric regression, *The Annals of Statistics*, **13**, 970-983.

[37] Stegmann M. B., Ersboll B. K., and Larsen R. (2003). FAME - A Flexible Appearance Modelling Environment. IEEE Trans. on Medical Imaging, 22(10):1319-1331.

[38] Trouvé, A. (1998). Diffeomorphism groups and pattern matching in image analysis. *Int. J. of Comp. Vis.*, 28(3):213–221.

[39] Turlach, B. A. (1997), Constrained Smoothing Splines Revisited, *Statistics research report SR 008-97, Australian National University.*

[40] Utreras, F.I. (1985). Smoothing noisy data under monotonicity constraints: Existence, characterization and convergence rates, Numerische Mathematik (47) 611?625.

[41] van der Vaart, A. W. (1998). Asymptotic statistics. *Cambridge University Press*

[42] Wahba, G. (1990). Spline models for observational data, *SIAM, Philadelphia PA.*

[43] Younes, L. (2003). Invariance, déformations et reconnaissance de formes. *Mathématiques et Applications. Springer-Verlag.*

[44] Younes, L. (2004). Deformation analysis for shape and image processing, *Lecture notes.*